Introduction

In a recent issue of this journal an author named Connolly (2008) presented what he termed a ‘critical review’ of some of our previous work on the relative attainment of male and female students in UK schools. He proposed three general areas for criticism – our use of attainment gaps, our consideration of outcomes other than at specific thresholds, and our querying of the idea of student ‘underachievement’. These problems, he claimed, have ‘given rise to a number of misleading conclusions that have questionable implications for practice’ (p.250). However, those of his ‘criticisms’ with any merit are actually the same as our own conclusions, transmuted by Connolly from our papers that he cites, while his remaining ‘criticisms’ are based on faulty elementary logic. In case readers have not read our work and were somehow misled by Connolly, we give here a brief reply to each criticism in turn. This matters, because a greater understanding of patterns of attainment and of the nature of underachievement is a precursor to the design of successful initiatives to overcome inequalities in educational opportunity and reward. This is both a practical and an ethical issue.

Achievement gaps

Following Arnot et al. (1996, appendix), the idea of an achievement gap was used by Gorard et al (1999a) to illustrate the proportionate difference in public examination attainment between males and females, after adjusting for differences in the numbers of each sex entering any examination. Thus, the overall achievement gap is composed
of the attainment gap (the proportionate male:female difference in numbers attaining a
certain grade) minus the entry gap (the proportionate male:female difference in
numbers entering the assessment). Gorard et al. (2001, p.137) pointed out that entry
gaps between males and females in each A-level subject tend to be higher than they
are at GCSE. Substantially more female than male students enter A-level English each
year, for example, whereas the gap is less marked for entry to GCSE English. By
contrast, the eventual gap in attainment between male and females students is
generally lower at A-level than at GCSE. In English GCSE, female students attain
more of the high grades than males, but the attainment gap is much less at A-level.
One of the possible reasons for this difference is that the smaller proportion of each
age cohort of males who choose to continue with A-level English once the constraints
of the National Curriculum are over tend to be better at English. Thus, the entry gap
between sexes is greater but the attainment gap is then lower than at GCSE where
English was, and still largely is, compulsory.

Connolly (2008) objects to this analysis and, insofar as we understand it, what he
objects to is that it does not ‘properly disentangle the effects of differential entry’ and
differential outcomes. But this disentangling is precisely what the separate entry and
attainment gaps do. They can be examined in isolation (decomposed) for plausible
explanations of the entry gap itself, or the sex gap in eventual results. They can also
be combined to ‘control’ for entry in explaining achievement. All three are valuable,
and our body of work in this quite specific area of research has done all three.

Gorard et al. (2001, p.137) suggested that ‘in general, gender appears to play less of a
simply do not have the evidence to claim’ this. His given reason is that ‘there is some
evidence that assessment practices are significantly gendered at A level (see Murphy
and Elwood 1998)’. He is here committing the elementary logical error of confusing
‘less’ with ‘none’. Clearly (to us at least) assessment can be gendered at A-level, as
we and Murphy and Elwood (1998) published a decade ago, and still be less so than it
is at GCSE.

Connolly accuses Gorard et al. of ‘not comparing like with like’ (p.251) and tries to
explain this by saying:
A much smaller proportion of males tend to go on to take this subject [A-level English] than females. This, in turn, would suggest that the male entrants may well constitute a highly selective group that is more motivated and may also have achieved higher grades…’

It is difficult to see why this purported criticism by Connolly is not covered by many examples making precisely the same point in our own work. For example, the Gorard et al. (2001, p.131) paper, cited heavily by Connolly, says:

At present the only clear difference relating to gender in Mathematics is an entry gap of around 30% in favour of boys. If this were reduced it might lead to an achievement gap in favour of boys in its place (if it assumed that in general the strongest candidates are already more likely to be taking any subject)

Did Connolly miss passages such as this in the paper, despite citing it, or in our other work where the same point has been made repeatedly for years? Or could he not understand that the 2001 passage above, for example, makes the point he criticises us for not making? Or had he both read and understood the 2001 paper, but still decided to suggest to his readers that passages like the one above were not there, so that he could make a comment that might then appear both original and somehow critical of our work? Whatever the truth of that, his criticism does not stand up. The achievement gap decomposed into entry and attainment components can be, and has been, used for all of the things that Connolly wants. Our suggested conclusion, that he takes exception to, stands. It really does appear that as far as the assessment is concerned (once entry patterns have been accounted for), there is less of a difference between sexes at A-level than at GCSE. In the absence of evidence to the contrary, parsimony rules. The key difference lies in the prior entry patterns not the attainment of those sitting the examination. And the former is where we have directed future explanatory research - just as Connolly now also proposes but confusingly in a way that sounds as though it is opposition to us. Of course, no one dataset, analysis, index of attainment or paper (such as Gorard et al. 2001) is complete in itself. Whatever one does in one of these is always open to the comment that one could have done
something else or something further. This is a lazy comment and, in itself, is no criticism of the work that has been done.

**Different levels of analysis**

The second area of concern for Connolly is that in some of the pieces he cites, we have investigated achievement gaps at every level of outcome for any assessment. He wonders ‘whether such additional detail is actually necessary’ (p.251), and is worried that there are key thresholds, such as grade C at GCSE, which it is strategically important to surpass. As with his comments about the entry, attainment and achievement gaps (see above), looking at the gaps at other grades than C is not a problem at all. Analysing at each grade or level, allows easy aggregation into benchmarks such as grade C or above while also providing those who want it with greater insight (such as the previously unpublished and unremarked finding that achievement gaps do not exist at the lowest levels of attainment but increase with every grade until A* at GCSE). Gorard et al. (2001) deliberately reminded readers that a G grade, while often overlooked by political commentators, is the actual pass grade for GCSE and so presents another important threshold. Nevertheless, readers of Gorard et al. (2001) will see that much more of the analysis presented concerns the grade C boundary at GCSE than grade G. In the first subject covered (Mathematics) grade G is not even mentioned. That paper concludes with a more general consideration of the official benchmark figures at grade C and above. To suggest, as Connolly does (p.251) that we are being misleading in that we ‘imply that they [G grades] are all regarded popularly and in practice as passes’ is itself a distortion. The paper makes no such claim about popularity, and quite clearly states that in practice the C grade is the one that is considered relevant officially.

On the other hand, despite Connolly adopting without demur the government and media line about C being the real pass grade, it is our experience of field work in schools and colleges that many students are aiming for and rightly proud of genuine achievements at GCSE between D and G grades. A high proportion of achievement is in this band and we really cannot see any reason not to analyse it in addition to the A* to C grades. When Gorard (1998) and Gorard et al. (1999a) did so no one had done
the same kind of analysis with national data at either level. To do the analysis at both levels might seem advantageous. Connolly seems to be clutching at straws for criticism here. When he says ‘the finding that there is no gender gap at grade G or above is of very little practical use’ (p.251) he portrays a rather disappointingly elitist view of student performance.

At the time of writing the 2001 paper, and as shown in the associated literature reviews from the same study (e.g. Salisbury et al. 1999), the problem of boys’ underachievement was considered by others (including many of those cited approvingly by Connolly) to be a problem at low levels of attainment concentrated among those from the poorest SES background – the so-called failing boys. What Gorard et al. (2001) suggested (see above) was that this version of the problem was unlikely. Since the same proportion of males and females were gaining grade G or above in all GCSE subjects the problem was not one at the lowest levels. In fact, the problem was highest at the highest grades (at A* in particular). Connolly objects to our suggestion that it is, therefore, ‘highly unlikely the gender gap is primarily to do with disaffected boys from economically depressed families’. He objects despite citing his own work which ‘shows clearly that while the gender gap appears to be relatively small it is actually constant across all social groups’. He objects because he somehow believes that these are ‘very different conclusions to those arrive [sic] at by Gorard et al.’ (p.253). Again, he seems to be making an elementary error of logic. Gorard et al. (2001) argued that the gender gap is not primarily to do with low SES boys, which seems entirely consistent with Connolly’s subsequent claim that the gap is constant across social groups. Let us make that simple error more explicit for Connolly. If the achievement gap is constant across social groups then it cannot be primarily to do with low SES boys. This also seems obvious (to us at least).

**What is underachievement?**

The third area of unfounded criticism by Connolly relates to the purported problem of our understanding and operationalising of the idea of ‘underachievement’. Again, it is important to set the context for this work. The impetus for the research stemmed from the large amount of academic, professional and popular writing which, in the late
1990s, was concerned with 'crisis' accounts of falling examination standards and of underachievement in school. These accounts related particularly, but not exclusively, to working class boys at secondary level. The suggested remedies, often enacted without evidence of efficacy, included single-sex teaching and alternate-sex seating in class. We were particularly concerned with the potential confusion and conflation between low achievement (in absolute terms), lower achievement (relative to other individuals or groups) and underachievement itself (not achieving as much as the best prediction available). These three things are very different, both analytically and substantively.

Smith (2003a, 2003b) based a model of underachievement on the definition by Thorndike (1963, p.19) as 'achievement falling below what would be forecast from our most informed and accurate prediction, based on a team of predictor variables'. The work was in secondary schools because it was here that most concern was expressed at the time, and many of the strategies to ameliorate the 'underachievement of boys', such as single sex teaching and mentoring, had been targeted. The study involved the collection of a large amount of contextual information from official statistics and via fieldwork in the case schools, which previous research had suggested might have an impact on academic attainment (for example, family background, motivation, prior attainment, sex and eligibility for free school meals). Smith (2003a, 2003b) then created and carefully explained models of KS2 to KS3 progress to see the extent to which boys (and others) could be said to be underachieving in relation to the best contextualised value-added prediction. Once potential under-achievers had been identified, a series of focus group discussions were undertaken in order to try and tease out any additional features which might characterise those who, according to the statistical model, could be reliably identified as 'underachievers' and 'overachievers' (Smith 2007).

What Smith (2003a, 2003b) shows is that once prior attainment and other contextual factors are controlled for there is little variation in attainment at KS3 to be explained by the sex of the student. The young people identified as underachieving in this model were a very heterogeneous group comprising both boys and girls, and including those with both high and low levels of attainment in absolute terms. Smith (2005), and elsewhere, is able to describe the modelling in greater detail and demonstrates that
many of the ‘explanatory’ variables are interchangeable in the model. Omitting prior attainment from the model leads to more variance linked to the contextual factors including sex, and vice versa. We do not need Connolly to explain that if prior attainment scores or context values are omitted from the models then the variance left to be explained by sex (or indeed any other factor) increases. We have already published this many times, and in many areas of investigation. It is one of the reasons why Gorard (2006, p.82), as a recent example, warns that regression is not a test of anything – ‘A statistical result, such as provided by regression, is only the start of an investigation, not its end’. Hence the focus groups and subsequent work conducted by Smith (2007). Unfortunately, Connolly (2008) does not seem to have read our wider work in this area, and most notably omits reference to Smith (2005), and to the fullest treatment in Smith (2002).

Connolly says:

All that Smith can validly claim from the findings of her analysis is that gender can only explain 0.8% of the variation in attainment above and beyond the effects it has already had up until the age of 11. (p.254, emphasis in original)

This is correct. And since this is all that Smith does claim there is no problem, as Connolly should know even from reading Smith (2003a, 2003b). What the analysis suggests is that activities and initiatives based on overcoming the gendered effects of early secondary education are unwarranted. What the fuller picture of our work shows is that the gender gap in attainment appears very early in the lifecourse, but does not worsen substantially from KS2 to KS3. We suspect that Connolly agrees with us here, even though he again phrases this agreement as a ‘criticism’. The problem with his suggestion of taking the longest view possible (such as provided by a cohort study) is that we sacrifice completeness and validity of measures for such a longitudinal approach. The ‘narrow’ range of KS2 to KS3 means that Smith (2003b) was working with a higher quality dataset than usually provided by a cohort study in which dropout and the difficulty of assessing attainment among pre-school children are inevitable. Neither approach is clearly preferable. And, in our own analyses we have done both many times, and both approaches seem to show the same substantive results (Gorard et al. 1999b, Gorard and Rees 2002, Gorard et al. 2007). The findings are that
inequalities in attainment patterned by SES and related variables start early in life and that secondary school events do little to remedy them (Gorard et al. 1999c, Gorard et al. 2003, Gorard and Selwyn 2005, Gorard and See 2008).

Connolly (2008) also seems to object to consideration of the amount of variation in attainment scores between Key Stage 2 and 3 that can be accounted for by variables such as sex and eligibility for free school meals. He writes: 'what Smith's model is effectively doing is controlling for the effects of poverty on educational attainment up to the age of 11 and then looking at what additional effects above and beyond this does poverty have from the ages of 11 to 14' (p.254). This is true but far from objectionable. Our interest was in the period of early secondary schooling (see above), and the results suggest that variables such as sex and even FSM have limited impact on attainment between these Key Stages. It seems unfair of Connolly not to acknowledge that we have already considered this issue. See, for example, Smith (2005, p.113):

The use of these variables in the construction of a model of underachievement is problematic… Any disadvantage the individual might experience prior to, or at the time of this assessment (and which could manifest itself in low achievement or underachievement) might effect [sic] these results and so bias the model (Smith 2005, p.113).

It also seems unfair of Connolly not to acknowledge that we have already described running several versions of the model, and that we had both a theoretical and pragmatic justification for retaining the model as it was. One approach we tried was to exclude prior attainment from the model altogether, relying on contextual characteristics alone to predict Key Stage 3 outcomes. Unsurprisingly the model was not as 'strong' as one which included a measure of prior attainment, and so it does not meet Thorndike’s definition of underachievement (see above). Another approach was to disregard the potential for bias that social, or other, disadvantage might have on prior attainment scores and to include these in the model without adjustment.

This could be justified on the grounds that the focus of the study is on identifying underachievement in the secondary school, and realistically
secondary schools can do little to prevent underachievement that might have taken place in the primary school: their influence on any underachievement can only be remedial in response to the needs of their intake (Smith 2005, p.113).

The decision was made to adopt the [second] of these two possible strategies: prior attainment was retained in the model but with its potential for bias noted (Smith 2005, p.113).

Clearly, a little fuller reading of texts published by us for a full six years or more before his paper would have allowed Connolly to see that his concern that we ought to 'remove other proxy measures of poverty from the analysis, most notably prior attainment' (p. 254), was unfounded. For a male professor of gender and education to so denigrate the work of a female PhD student without bothering to read either the thesis itself (Smith 2002) or the book of the thesis (Smith 2005) is unprofessional, and perhaps even unethical. We had already as a matter of course conducted the same analyses that he now suggests we conduct (as though this were some kind of criticism of our work). The model he cites was merely the most parsimonious developed and the best placed to address the prior research questions (see Smith 2002 for a fuller discussion).

Connolly also misunderstands and therefore seems to object strongly to the following statement:

We might say that a particular social group exhibits lower achievement… than another, as in the case of some ethnic groups… [However] this is very far from saying that the lower attaining group could and should do better on that assessment. (Gorard and Smith 2004, p.217, emphasis in original).

He states that this passage ‘is only one step away from suggesting that there is something inherent and/or natural about the poor educational attainment of particular minority ethnic groups’ (Connolly 2008, p.256). This is both ridiculous and very worrying (in terms either of his motive or his ability to read text). We assume that Connolly knows about the long-term and widespread literature on cultures and
registers in education, including the work of Basil Bernstein. But let us put the basic idea very simply. Imagine a student who speaks only English sitting a test written in Norwegian. The student might gain zero marks. Aggregating the figures we might find that English speakers as a group did considerably less well on the test than Norwegian speakers. Therefore, we might correctly say that this particular group ‘exhibits lower achievement’ but would not say that the English speakers ‘could and should do better on that assessment’ (emphasis in original, see above). Our condemnation here, and in Gorard and Smith (2004), is quite clearly of the inappropriateness of the assessment for this student or group of students, rather than of some inherent lack of talent among the students involved. For more on this see, for example, Gorard (2004).

Connolly (2008) criticises us for being unaware of the basic distinction between sex and gender and for our purported lack of engagement with the ‘basic literature on gender’ (p.256). He says ‘It would seem that Gorard and Smith are unaware of the basic distinction between ‘sex’ and ‘gender’ made by Oakley (1972)…..’ (p.256), because we suggest that a basic practical problem with identifying sex as a factor in attainment is that we cannot change the sex of the student (usually) and so to make practical progress we need to focus on identifying the factors that educators can change (and with which sex might interact).

He bases this judgement about us being unaware of ‘gender’ on another of his errors of logic, and his criticism here fails on at least three counts. First, there are many things that we are aware of that are not mentioned in the Gorard and Smith (2004) paper. Lack of citation is not evidence of lack of awareness. Obviously so, we would have thought. Is Connolly unaware of the theory of relativity, for example, if he has never discussed it in his writing? However, if Connolly had really read any of our literature reviews that he cites, such as Salisbury et al. (1999), he would have found the literature he sought. If he wishes to base an argument on absence of something in our writing then he has a clear responsibility to search exhaustively, in a way that he clearly has not. The fault again seems to lie not with us but with his limited reading of research. Second, although we are aware of this distinction and know Oakley and admire much of her work, we do not find this particular distinction very useful here, or believable in general. If Connolly really believes that gender, as he envisages it,
could have been operationalised and used in this study as well as sex, then we invite him to propose how. As with any concept or theory it is not illogical not to believe it or not to use it. Third, the supposed distinction is not relevant here anyway. The analysis was of attainment by sex, as defined by the identifying characteristics of the test candidates in official statistics. Gorard and Smith (2004, p.219) said:

Even if we could show that sex, for example, was a cause of a level of achievement, we could not adjust the sex of individuals to ameliorate low achievement.

What this shows is that sex might be one useful way of classifying students and so discovering an inequality in attainment but that we must look wider than sex itself for an explanation and so a remedy. Part of this wider picture is what Connolly explains, rather poorly and patronisingly, as ‘gender’ rather than sex and so he agrees with us (again) while appearing not to.

Connolly continues by claiming that when Gorard and Smith (2004, pp.204-205) make a suggestion that follows clearly from the proposed problem of unfair assessment (above) this portrays our lack of wider knowledge of research in this area. We point out that where there are surface inequalities in test scores between social groups then we could change the nature of the assessment ‘in the same way that ‘neutral’ IQ tests were developed, if that is what is desired’. Connolly (p.257) says that it is ‘particularly remarkable – and ultimately irresponsible – that they could make such a suggestion’ because it purportedly ignores prior feminist critique of a completely different approach to the one we discuss. Connolly quite wrongly believes that changing the nature of (or calibrating) the assessment is ‘precisely’ the same thing as ‘the approach popularly advocated and taken during the 1990s in terms of ‘masculinising’ the curriculum and placing more emphasis on male role-models in an attempt to make education more appealing to boys’. The apparently unsuccessful attempts based on changing the curriculum or attracting more male teachers are nowhere near ‘precisely’ the same approach as that of calibrating the tests to be sex-neutral. It is not us that is being irresponsible here. It seems irresponsible of Connolly to suggest some lack on our part in not citing a literature that is not relevant (also see above).
A more substantive and interesting area of difference appears on p.257 of Connolly (2008). Gorard and Smith (2004) cite the sudden appearance of the current much larger gap in attainment between male and female at age 16 at the same time as the reforms of the late 1980s such as GCSE, abolition of norm-referencing, more coursework and so on as suggestive of an explanation. We reproduce one part of this analysis in Figure 1, representing the GCSE/GCE results of all school-leavers in England from 1974/75 to 1997/98, separated by male and female. For details of the calculation involved see Gorard (2001). What this shows is the abruptness of the main increase in the achievement gap at that time, and the otherwise volatile and small-scale nature of change over time. Therefore, what is needed to explain the 1987/88 change is a cause that is one-off and immediate in impact. Connolly objects because this leaves little room for his favoured explanations such as equal opportunities legislation and the gradually increasing impact of feminism on teachers’ practice. We think he is wrong here, in not allowing evidence to disturb a cherished notion (although we are sympathetic to the problems that may cause him and his prior work in the area). But he is not alone in that, and a fuller discussion on this important issue (see, for example, Gorard 2006b) would have been more interesting for readers than the insubstantial points Connolly uses most of his paper for.

Figure 1 - Achievement gap in favour of females attaining 5+ GCSE A*-C
Conclusions

Connolly’s piece appears to us as largely argument for argument’s sake. The purpose of his review is unclear, as all of his sensible concerns are already clearly presented in our prior work. He laments that we do not engage more with the theories and concepts in the field of gender and education. But he himself does not cite the book Smith (2005) which is replete with references, nor the more detailed treatment in Smith (2002), as cited in Smith (2003a, 2003b). He does cite our ‘State of the Art’ review (Salisbury et al. 1999), and so presumably knows that we there select, review and reference over 150 pieces from this very literature, and use it as the basis for our subsequent analysis of official data. Only reasons of space prevented us from using more of the prior literature, such as appears in our subsequent writings in this area.

Connolly does end his paper well, saying that ‘we cannot allow quantitative research simply to be left to statisticians. Good quantitative research is never simply about
numbers…’ (p.258). We agree, since this is what we have already written about regularly (e.g. Gorard 2006c), and for a considerable period (e.g. Gorard 1997). So Connolly is merely, and unfairly, repeating our own words as though they were a criticism of us. Neither Smith nor Gorard is a statistician, and neither of us are what Connolly would term ‘quantitative’ researchers. Again, it would be relatively easy for Connolly to have followed our references in the literature that he does cite to discover us saying in many ways though many outlets for many years that good quantitative research is never simply about numbers (e.g. Gorard with Taylor 2004), and to see this illustrated in our work with numbers (e.g. Gorard 2008). To make it sound as though this were a new idea and somehow opposed to our practice is simply dishonest. In fact his whole paper relies for any effect on readers, like Connolly and his reviewers, not looking closely at the analyses and literature referred to and so coming away with the erroneous impression that some kind of technical or theoretical debate is involved.

References


Gorard, S., Salisbury, J. and Rees, G. (1999a) Reappraising the apparent underachievement of boys at school, Gender and Education, 11, 4, 441-454


